Hierarchical Image-Guided 3D Point Cloud Segmentation in Industrial Scenes via Multi-View Bayesian Fusion

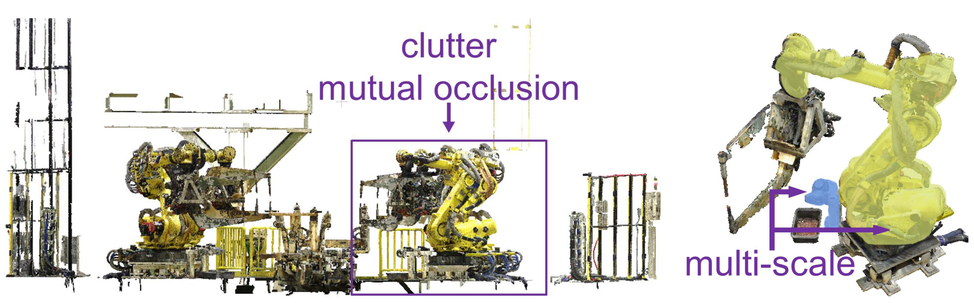
Yu Zhu[1], Naoya Chiba[2], and Koichi Hashimoto[1]

[1] Intelligent Control Systems Lab, Department of System Information Sciences, Graduate School of Information Sciences, Tohoku University

[2] D3 Center, The University of Osaka

Introduction

Goal: Part-level 3D segmentation for large, cluttered industrial scenes.



Issues:

- 3D annotation costly, no public dataset
- Mutual occlusion, cluttered, multi-scale between objects
- semantic inconsistency across views

Our solutions:

- Image-guided, easy for training
- Hierarchical structure to decouple clutter, occlusion, and multi-view
- Bayesian fusion enforces consistency

Methodology

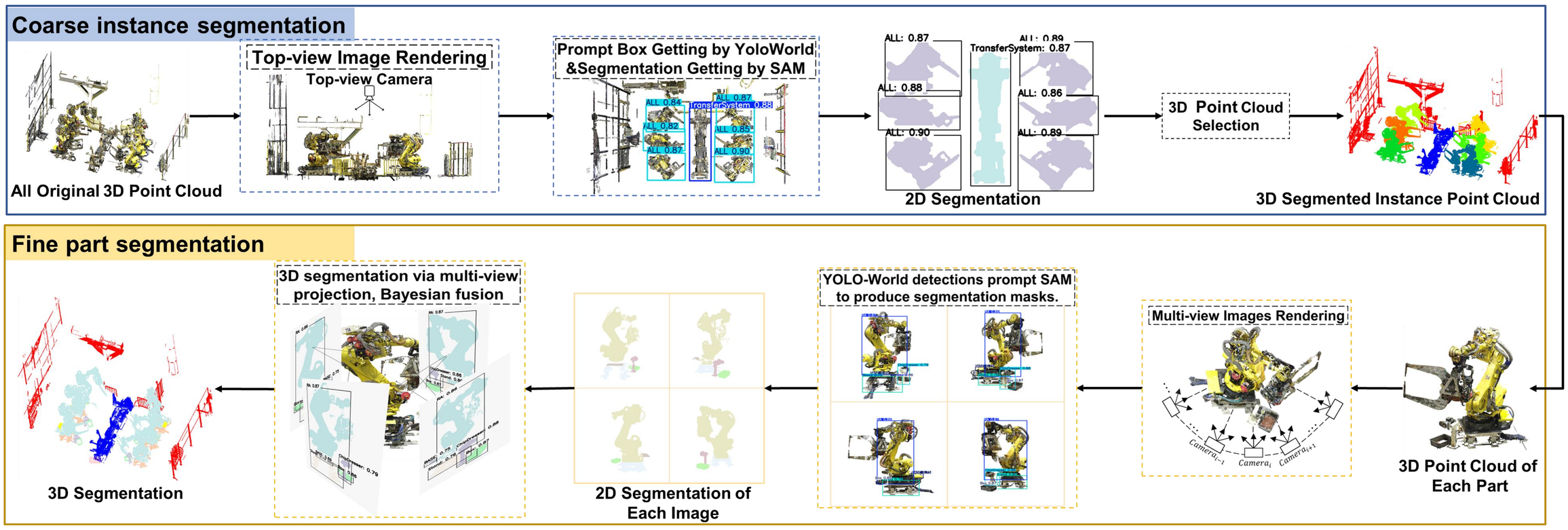


Figure 1: The pipeline of the proposed hierarchical 3D segmentation system. Two main stages: (1) instance-level segmentation; (2) part level segmentation.

Hierarchical structure:

- **Instance**: extract instances first to decouple clutter/occlusion;
  ↓
- **Part**: segment parts within each instance, giving clearer boundaries & cross-view consistency

- **Detection:** YOLOWorld prompt SAM, cutting annotation cost
  ↓
- **Segmentation**: boundary-precise, no extra training required.

Instance level via top-view projection: It provides wider coverage and minimizes self-occlusion, enabling fast, coarse instance generation as reliable priors for part-level refinement.

Part level multi-view consistency via Bayesian Fusion: For each instance and view, back-project the mask to 3D and match pixels to the nearest raw point with a depth-consistency check. To enforce cross-view consistency, we update the matched 3D point X's posterior given the per-view observation x^{θ} via Bayesian fusion:

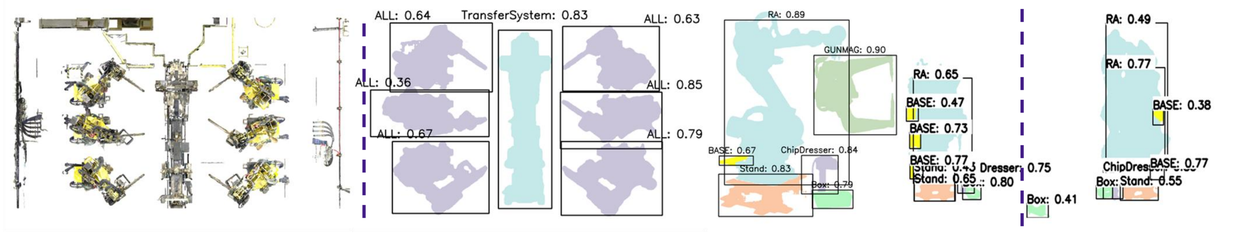
$$p_{\theta}(c \mid X) = \frac{\alpha_{\theta}}{Z_{\theta}}\, p\left(x^{\theta} \mid c\right) p_{\theta-1}(c \mid X)\ ,\ \ \alpha_{\theta} = \frac{1}{3}(A + N + B).$$

Each view is weighted by three quality factors—A: reasonable mask size, N: sufficient in-mask 3D points, and B: clear, coherent boundaries—so higher-quality views receive greater weight in Bayesian fusion.

Experimental Results

We evaluate on factory data and a public dataset. In factories we fine-tune two YOLO-World detectors (instance/part)—light labels, minutes of training; on the public dataset we compose single objects into multi-object scenes and run our pipeline without extra training, showing generality and zero-shot transfer.

Industrial Dataset

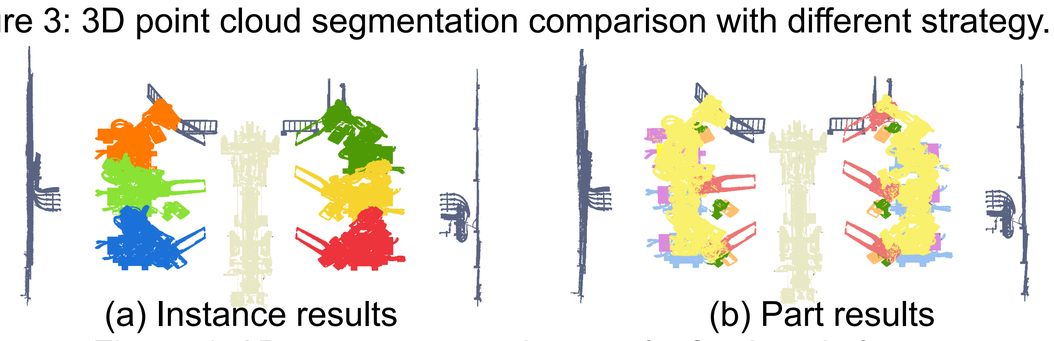


(a) Top-view image (b) Based on our hierarchical framework (c) Single-stage

Figure2: 2D segmentation comparison

(a) Projection-only (b) Cluster (c) Bayes (d) GT

Figure 3: 3D point cloud segmentation comparison with different strategy.



(a) Instance results (b) Part results

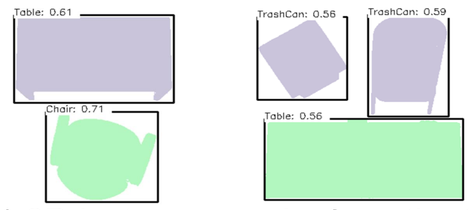
Figure 4: 3D part segmentation results for the whole scene.

Objects	Precision			Recall			mIoU		
	Bayes	Cluster	Original	Bayes	Cluster	Original	Bayes	Cluster	Original
RA	**67.42**	56.31	54.67	**98.03**	94.13	94.25	**65.29**	53.84	52.67
Dresser	**98.96**	86.31	67.94	**97.28**	88.24	88.97	**96.07**	73.91	58.73
Box	**99.03**	97.84	89.26	**99.12**	78.52	78.84	**98.01**	75.43	68.06
Stand	**87.69**	78.50	71.36	**97.23**	86.20	86.37	**83.43**	68.06	63.93
Base	**96.24**	86.83	83.92	**99.31**	75.84	75.29	**96.01**	66.67	63.93
GUN	**98.24**	79.66	59.32	67.89	**93.00**	92.14	57.48	**72.43**	55.03

Table 1: Per-category comparison across each segmentation strategy

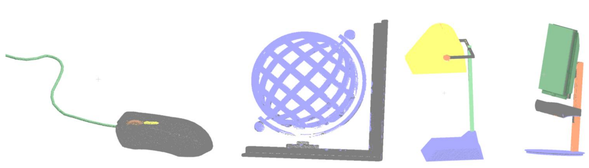
PartNet Dataset

(a) Composite scene



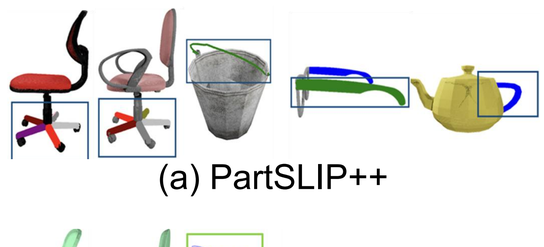
(b) 2D instance masks from the top view

(c) 3D instance segmentation



(d) Some results of part segmentation

Figure 5: Some part-segmentation results on composed scenes.



(a) PartSLIP++

(b) Our method

Figure 6: Part-level segmentation comparison with PartSLIP++

Category	PointNet++	PartSLIP	PartSLIP++	Ours
TrashCan	71.73	22.30	22.93	**73.09**
Keyboard	74.50	**72.85**	72.35	68.42
Globe	46.50	92.80	**96.50**	81.19
FoldingChair	10.90	83.60	89.90	**91.69**
Mouse	12.10	44.03	46.60	**47.08**
Scissors	49.96	60.33	**60.46**	52.54
Bottle	48.80	83.40	85.85	**87.62**
Knife	35.40	65.20	64.30	**70.34**
Cart	11.60	88.10	84.40	**88.63**
Dispenser	12.10	73.70	72.00	**73.81**
Suitcase	40.60	64.50	70.00	**71.99**
KitchenPot	31.70	69.60	**76.80**	76.40
Lamp	68.00	66.30	67.90	**68.56**
Door	45.70	43.10	45.10	**56.53**
Table	**63.70**	44.30	45.30	44.93
Kettle	20.90	77.00	**85.60**	75.96
Chair	42.20	85.30	**85.30**	78.71
Bucket	0.00	**89.60**	85.50	86.46
Eyeglasses	78.65	77.00	88.25	**89.38**
Display	30.02	**84.80**	81.80	68.54
Camera	9.70	61.80	**63.20**	61.31
Furniture	46.90	53.60	57.20	**62.78**
Oven	31.70	69.60	70.30	**72.96**
Overall	38.41	68.38	70.33	**71.70**

Table 2: Per-category segmentation mIoU(%) across 23 categories.